Audio Quality Research

Outputs

- Technical publications and presentations demonstrating new research results.
- Algorithms and data supporting speech and audio coding and quality assessment.
- Objective estimates and subjective measurements of speech and audio quality.

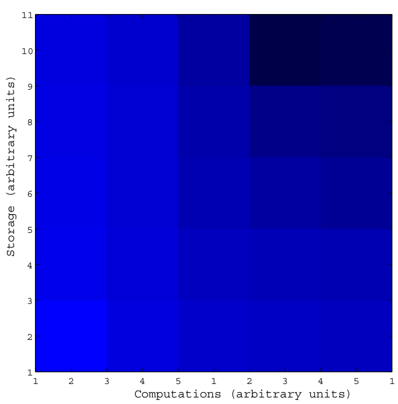

Figure 1. Bit-rate as a function of codec computations and storage. Higher rates are brighter blue.

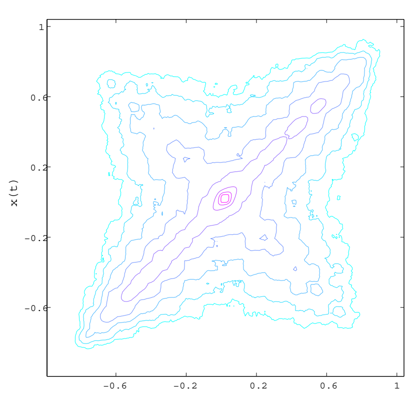

Figure 2. Contours of constant relative frequency for two-dimensional histogram of adjacent speech samples.

Digital coding and transmission of speech and audio signals are enabling technologies for many telecommunications and broadcasting services including cellular telephone services, voice over Internet protocol (VoIP) services, and digital audio broadcasting systems. Speech and audio signals can now be coded and transmitted at remarkably low bit-rates with good fidelity. In addition, coded speech and audio signals can be packetized for transmission, thus sharing radio spectrum or wired network bandwidth with other data streams and hence with other users.

In digital coding and transmission there is a four-way trade-off among quality, bit-rate, delay, and complexity. The ITS Audio Quality Research Program works to identify and develop new techniques to increase quality or lower the bit-rate, delay, or complexity of digital speech and audio coding and transmission. The ultimate result of such advances is better sounding, more reliable, more efficient telecommunications and broadcasting services.

In one FY 2004 Program effort, a family of speech coders with fixed speech quality and fixed delay was developed. For this family, the four-way trade-off described above becomes a 2-way trade-off: complexity vs. bit-rate. Complexity is comprised of two major factors: a computational requirement and a storage requirement. Figure 1 shows how bit-rate can be reduced by increasing either computations or storage. For a given implementation, one would select the member of this speech coder family that gives the lowest bit rate and has computation and storage requirements consistent with the implementation platform.

The robustness of digital coding and transmission algorithms is critical in applications that use lossy channels such as those associated with wireless systems and those provided by the Internet. The Program has also continued to work towards more robust speech coding through a method called multi-descriptive coding (MDC). In MDC an encoder forms multiple partial descriptions of a speech signal and these descriptions are sent over different channels. If all descriptions arrive at the decoder intact, a higher-quality reconstruction of the speech is possible. If channel failures cause any of the descriptions to be lost, then a lower-quality reconstruction of the speech signal is still possible.

One MDC approach recently developed in the Program exploits the naturally occurring correlations between adjacent samples of speech. This correlation can be seen in the way probability mass is organized along the diagonals in the two-dimensional histogram in Figure 2. Working with these correlations, one can effectively decompose a stream of speech samples into two streams that provide maximum fidelity both individually and when combined.

In digital speech and audio systems, a set of complex time-varying interactions among signal content, source coding, channel coding, and channel conditions often make it difficult to define or measure speech or audio quality. The Audio Quality Research Program operates a subjective testing facility and runs controlled experiments to gather listeners' opinions of the speech or audio quality of various coding and transmission systems. The Program has also developed and verified tools for the objective estimation of telephone bandwidth speech quality.

In FY 2004 the tools for objective estimation were further enhanced through the development, optimization, and testing of a robust technique for tracking variable transmission delay for a wide range of speech coding conditions. This is important because transmission delay can vary significantly and rapidly in packetized speech transmission systems, even within a single spoken phrase. This delay variation arises from the basic nature of packetized data networks and can be mitigated, but not eliminated, through buffering techniques. To understand the resulting speech quality, it is imperative that this continually changing delay be accurately tracked.

VoIP is presently the most prominent example of a packetized speech transmission system, but some land-mobile radio systems are also now interconnected through packetized links. Thus the Program developed a tool that is effective across the range of VoIP conditions and land-mobile radio conditions. This was accomplished through detailed analysis of a broad database that includes over 6000 different recordings. This analysis led to a successive refinement approach that can respond appropriately to each of the different conditions of interest. Figure 3 provides an example of actual and estimated delay histories for a worst-case condition involving 2.45 kbps speech coding and a high rate of packet impairments.

Signal content plays an important role in speech quality and thus the type and amount of background noise that is combined with speech can have a significant effect on speech quality. As cellular phones have become ubiquitous, they are routinely used in locations with significant background noise. In FY 2004 Program staff developed a database of digital recordings of background noise signals. This database includes recordings from bus and car interiors, office, coffee shop, and party environments, and urban sidewalk environments as well. The recordings are used in the Program to create realistic and diverse environments for speech quality assessment.

Throughout FY 2004, subjective and objective audio quality testing was conducted to support Audio Quality Research efforts. In addition, Program staff continued with selective upgrades to the ITS Audio-Visual Laboratories to keep them abreast of the state-of-the-art. Program staff continued to transfer technologies to industry, Government, and academia throughout FY 2004, using technical publications and lectures, laboratory demonstrations, and by completing peer reviews for technical journals and workshops. Program publications and other results are available at http://www.its.bldrdoc.gov/audio.

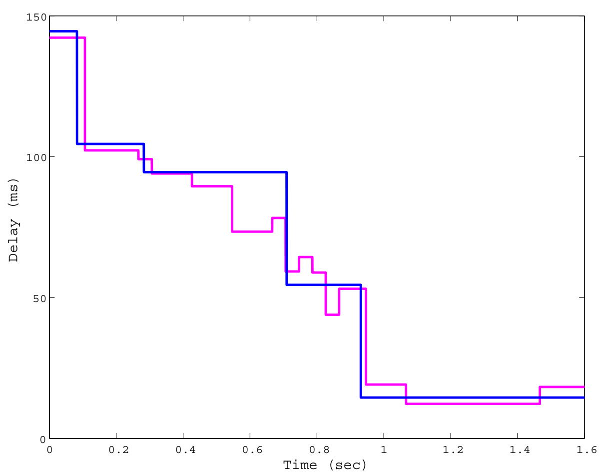

Figure 3. Actual (blue) and estimated (magenta) delay histories for 2.45 kbps speech codec with highly impaired packetized transmission.

Recent Publications

S.D. Voran, "Compensating for gain in objective quality estimation algorithms," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, May 2004.

S.D. Voran, "A bottom-up algorithm for estimating time-varying delays in coded speech," in *Proc. 3th International Conference on Measurement of Speech and Audio Quality in Networks (MESAQIN)*, Prague, Czech Republic, May 2004.

For more information, contact:
Stephen D. Voran
(303) 497-3839
e-mail svoran@its.bldrdoc.gov